

Hooking up Semantic MediaWiki with external tools via SPARQL

The Use case behind the RDFIO extension

RDFIO Extension: Google Summer of Code project in 2010

Mentor: **Denny Vrandečić** <<http://simia.net/>>

SMW + Bioclipse integration: MSc Thesis project in Bioclipse group, Inst. of Pharmaceutical Biosciences, Uppsala University, Sweden.

Supervisor: **Egon Willighagen** <<http://chem-bla-ics.blogspot.com>>

Samuel Lampa <<http://saml.rilspace.org>>, SMWCon 23 Sep 2011



This talk

- The motivation and use case behind RDFIO
- Looking for feedback on design decisions
- Outlook and Plans
- Possible collaborations / interactions?

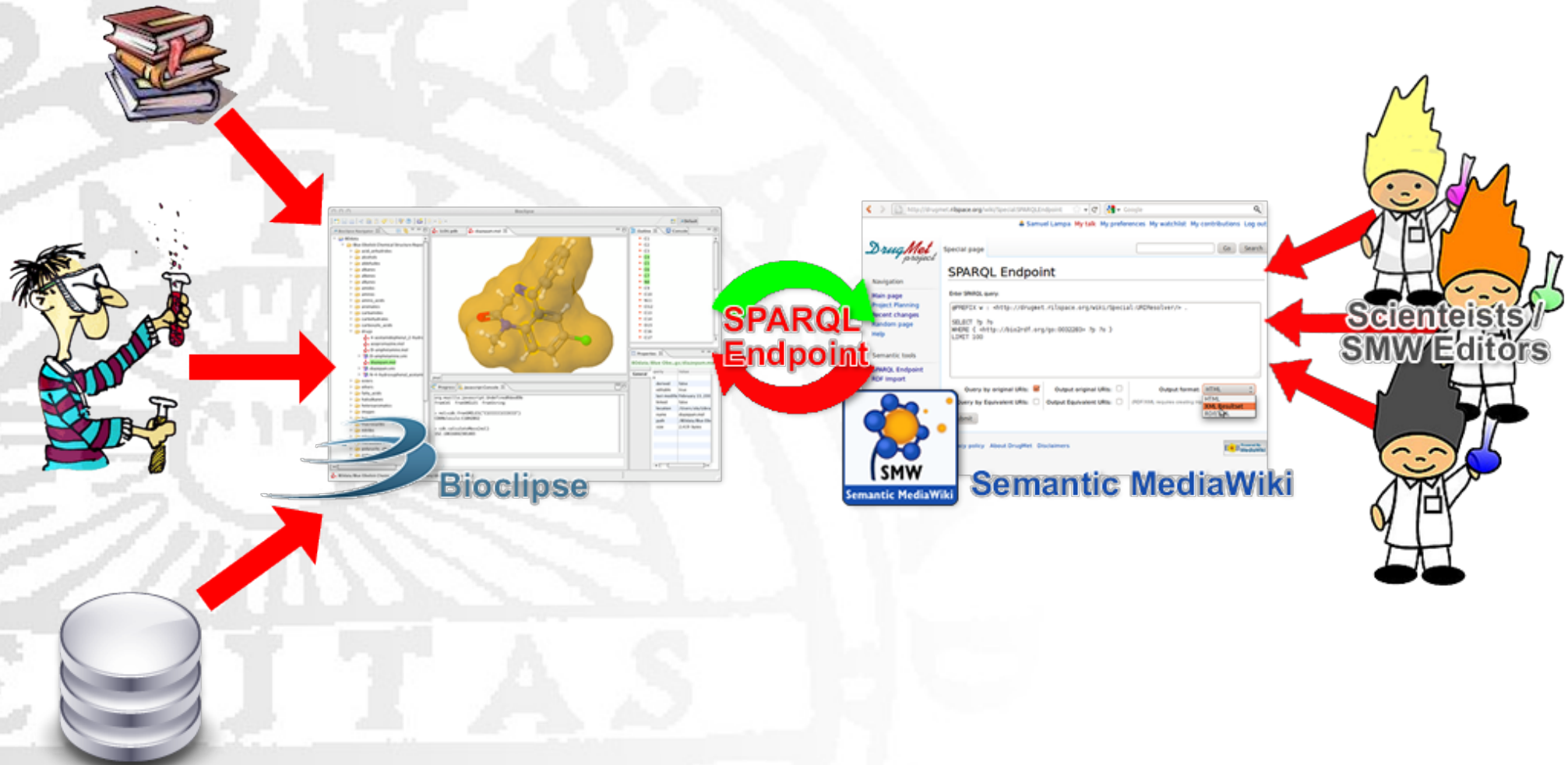


The Problem

- Need to combine:
 - Manual schema exploration, as knowledge grows
 - Automated data generation, for integrating data from literature or own experimental or computational results
 - Community collaboration
- Thus need something that supports all in same framework



Solution: Bioclipse + SMW



Bioclipse as automated or interactive batch writer



Bioclipse Bioinformatics Workbench

The screenshot displays the Bioclipse Bioinformatics Workbench interface. The central window shows a 3D molecular model of diazepam, a benzodiazepine, with its chemical structure overlaid on a yellow surface representation. The interface includes several panels:

- Bioclipse Navigator:** A tree view on the left showing a hierarchy of chemical structure repositories. The 'drugs' folder is expanded, and 'diazepam.mol' is selected.
- Outline:** A panel on the right listing the atoms of the selected molecule: C1, C2, C3, C4, C5, C6, C7, N8, C9, C10, N11, C12, C13, C14, O15, C16, and C17.
- Properties:** A panel on the right showing the general properties of the selected molecule, including its name, path, and size.
- JavaScript Console:** A panel at the bottom showing the execution of JavaScript code for molecule manipulation.

The Bioclipse logo is visible in the top left corner of the interface.

General Properties:

Property	Value
derived	false
editable	true
last modified	February 13, 200
linked	false
location	/Users/ola/Libra
name	diazepam.mol
path	/BOdata/Blue Obelisk
size	2,419 bytes

JavaScript Console:

```
org.mozilla.javascript.Undefined@dee89b
fromCml fromSMILES fromString

> mol=cdk.fromSMILES("C1CCCC1CCOCCO")
CDKMolecule:C10H20O2

> cdk.calculateMass(mol)
152.10616882301883
```



Bioclipse Shortfacts

- Open Source
- Based on Eclipse - Modular, Solid
- Cheminformatics/Bioinformatics Tools
 - Jmol, JChemPaint, etc
- Scripting (JavaScript)
- Semantic Web Support
- Core team at Uppsala University
- <http://bioclipse.net>



SMW Limitations

- No (prior) RDF import
- No obvious way to choose wiki titles for URIs
(Problem 1)
- RDF Export format not same as import format
... is hard-coded, using URIs such as:

`http://example.com/wiki/Special:URIResolver-3ASomeEntity`

(Problem 2)



So, needed a new extension:
RDFIO



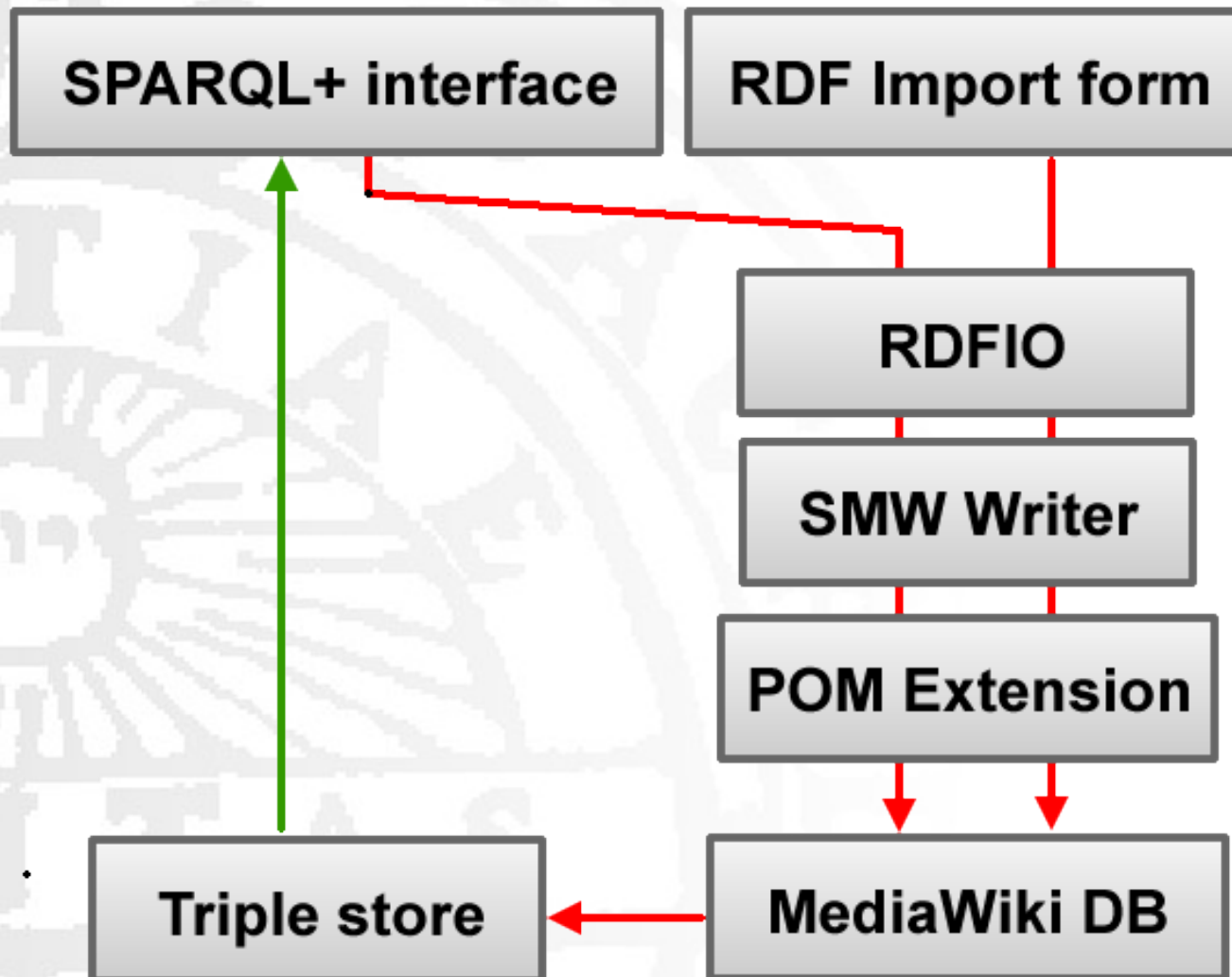
RDFIO Extension Short facts

- Google Summer of Code project 2010
 - Mentor: Denny Vrandecic
- Dependencies:
 - Semantic MediaWiki
 - SMWWriter (Denny Vrandecic)
 - POM (Sergey Chernychev)
 - ARC2 RDF Lib (Benjamin Nowack)
- Current version: 0.5
 - Needs bugfixing
 - Needs making SMW 1.6 compatible

● <http://www.mediawiki.org/wiki/Extension:RDFIO>



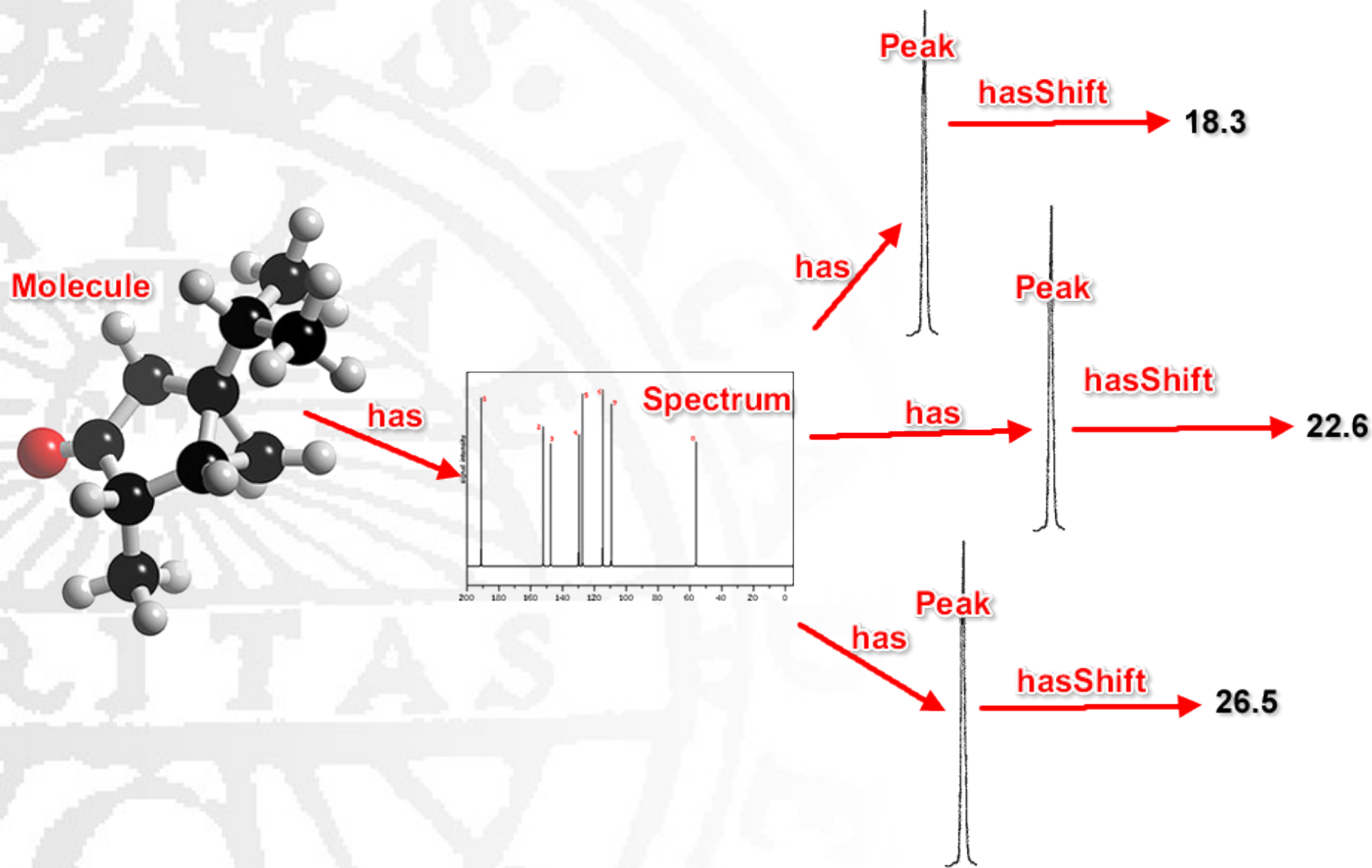
RDFIO Data Flow



Problem 1: Choosing wiki page titles for RDF entities



Example of Experimental Data



```
<http://[...]/nmrshiftdb/?moleculeId=234>
  dc:title "warburganal";
  chem:casnumber "62994-47-2";
  nmr:moleculeId "234";
  nmr:hasSpectrum <http://[...]/nmrshiftdb/?spectrumId=4735>;
<http://[...]/nmrshiftdb/?spectrumId=4735> nmr:field "50";
  nmr:hasPeak <http://[...]/nmrshiftdb/?s4735p0>,
    <http://[...]/nmrshiftdb/?s4735p1>,
    <http://[...]/nmrshiftdb/?s4735p2>,
    <http://[...]/nmrshiftdb/?s4735p3>;
  nmr:solvent "Chloroform-D1 (CDCl3)";
  nmr:spectrumId "4735";
  nmr:spectrumType "13C";
  nmr:temperature "298".
<http://[...]/nmrshiftdb/?s4735p1>
  nmr:hasShift 18.3;
  a nmr:peak.
<http://[...]/nmrshiftdb/?s4735p2>
  nmr:hasShift 22.6;
  a nmr:peak.
<http://[...]/nmrshiftdb/?s4735p3>
  nmr:hasShift 26.5;
  a nmr:peak.
```



Problem: URIs not nice as titles

Set \$wgLogo to the URL path to your own logo image.

navigation

- [Main page](#)
- [RDF Import](#)
- [SMW RDF Connector](#)
- [ARC2 Admin](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

search

toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Browse properties](#)

[Admin](#) [my talk](#) [my preferences](#) [my watchlist](#) [my contributions](#) [log out](#)

[page](#) [discussion](#) [edit](#) [history](#) [delete](#) [move](#) [protect](#) [watch](#) [refresh](#)

Http://pele.farmbio.uu.se/nmrshiftdb/:spectrumId4735

Facts about [Http://pele.farmbio.uu.se/nmrshiftdb/:spectrumId4735](#) [RDF feed](#)

Nmr:field 50 + 🔍

Nmr:hasPeak [Http://pele.farmbio.uu.se/nmrshiftdb:s4735p1](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p2](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p3](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p4](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p5](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p6](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p7](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p8](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p9](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p10](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p11](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p12](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p13](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p14](#) + 🔍,
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p15](#) + 🔍, and
[Http://pele.farmbio.uu.se/nmrshiftdb:s4735p0](#) + 🔍

Nmr:solvent [Chloroform-D1 \(CDCl3\)](#) + 🔍

Nmr:spectrumId [4735](#) + 🔍

Nmr:spectrumType [13C](#) + 🔍

Nmr:temperature [298](#) + 🔍



So, how to choose the wiki
page title then?



Let's look what we have in the
RDF, that might be used ...



Wiki page title for the Molecule?

```
<http://[...]/nmrshiftdb/?moleculeId=234>  
dc:title "warburganal";  
chem:casnumber "62994-47-2";  
nmr:moleculeId "234";  
nmr:hasSpectrum <http://[...]spectrumId=4735>;
```



Wiki page title for the Molecule?

```
<http://[...]/nmrshiftdb/?moleculeId=234>  
dc:title "warburganal";  
chem:casnumber "62994-47-2";  
nmr:moleculeId "234";  
nmr:hasSpectrum <http://[...]spectrumId=4735>;
```



Wiki page title for the Spectrum?

```
<http://[...]nmrshiftdb/?spectrumId=4735>  
nmr:field "50";  
nmr:hasPeak <http://[...]nmrshiftdb/?s4735p0>,  
    <http://[...]nmrshiftdb/?s4735p1>,  
    <http://[...]nmrshiftdb/?s4735p2>,  
    <http://[...]nmrshiftdb/?s4735p3>;  
nmr:solvent "Chloroform-D1 (CDC13)";  
nmr:spectrumId "4735";  
nmr:spectrumType "13C";  
nmr:temperature "298".
```



Wiki page title for the Spectrum?

```
<http://[...]nmrshiftdb/?spectrumId=4735>  
nmr:field "50";  
nmr:hasPeak <http://[...]nmrshiftdb/?s4735p0>,  
    <http://[...]nmrshiftdb/?s4735p1>,  
    <http://[...]nmrshiftdb/?s4735p2>,  
    <http://[...]nmrshiftdb/?s4735p3>;  
nmr:solvent "Chloroform-D1 (CDC13)";  
nmr:spectrumId "4735";  
nmr:spectrumType "13C";  
nmr:temperature "298".
```



Wiki page title for the Spectrum?

```
<http://pele.farmbio.uu.se/nmrshiftdb/?s4735p1>  
nmr:hasShift 18.3;  
a nmr:peak.
```



Wiki page title for the Spectrum?

```
<http://pele.farmbio.uu.se/nmrshiftdb/?s4735p1>  
nmr:hasShift 18.3;  
a nmr:peak.
```



So ...

- So, we might find useful titles in:
 - Dublin Core fields (small vocabulary for cataloging)
 - Custom identifiers
 - (parts of) URIs
- So, we need a flexible configuration system such as:
 - A system-wide defaults in **LocalSettings.php**, AND
 - **Interactive** configuration in **RDF Import form**

(See following slides ...)



In LocalSettings.php ...

```
/**
 * Used to generate "Pseudo-NameSpaces"
 */
$rdfiogBaseURIs = array(
    "http://example.org/someOntology#" => "ont1",
    "http://example.org/anotherOntology#" => "ont2"
);

/**
 * "Properties" to be used as Wiki Title
 */
$rdfiogPropertiesToUseAsWikiTitle = array(
    'http://semantic-mediawiki.org/swivt/1.0#page',
    'http://www.w3.org/2000/01/rdf-schema#label',
    'http://purl.org/dc/elements/1.1/title',
    'http://www.w3.org/2004/02/skos/core#preferredLabel',
    'http://xmlns.com/foaf/0.1/name',
    'http://www.nmrshiftdb.org/onto#spectrumId'
);
```



... and in RDF Import Screen (1/2)

Special page

RDF Import

RDF/XML data to import:

```
<rdf:RDF
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:q0pred="http://bio2rdf.org/go_resource:"
    xmlns:ns0pred="http://www.w3.org/2002/07/owl#">

    <rdf:Description rdfs:about="http://bio2rdf.org/go:0032283">
        <n0pred:accession>G0:0032283</n0pred:accession>
        <rdfs:label>plastid acetate CoA-transferase complex
[go:0032283]</rdfs:label>
    </rdf:Description>
</rdf:RDF>
```

Data format: RDF/XML [Paste example data] [Clear]

Options for properties	Options for non-properties
<input type="checkbox"/> Use namespace prefixes in wiki titles	<input type="checkbox"/> Use namespace prefixes in wiki titles
<input checked="" type="checkbox"/> Show abbreviation screen	<input checked="" type="checkbox"/> Show abbreviation screen

UPPSALA
UNIVERSITET

Privacy policy About DrugMet Disclaimers

... and in RDF Import Screen (2/2)

http://drugmet.rilspace.org/wiki/Special:RDFImport

DrugMet
project

Special page

RDF Import

Choose Prefix for unabbreviated property namespaces:

http://bio2rdf.org/bio2rdf_resource:	<input type="text"/>
http://purl.org/dc/terms/	<input type="text"/>

Note: These abbreviations will not be used if, for a property, there already exists a wiki article with the same name.

Entity (i.e: non-property) URIs lacking namespace prefix

http://bio2rdf.org/go:0009329
http://www.geneontology.org/GO.cite.shtml
http://purl.org/obo/owl/GO#GO_0032283

Hint: Namespace prefix configuration can be added as an attribute in the RDF tag below according to the configuration.

RDF/XML data to Import:

```
<rdf:RDF
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  >
```



Result: *Problem 1 solved!*
Page get sensible titles
(*See screenshot on next slide*)




Set \$wgLogo
to the URL
path to your
own logo
image.

navigation

- [Main page](#)
- [RDF Import](#)
- [SMW RDF Connector](#)
- [ARC2 Admin](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)


search

 127.0.0.1 [talk for this ip address](#) [log in / create account](#)













[page](#)[discussion](#)[edit](#)[history](#)

Spectrum:4735


Facts about Spectrum:4735 ⓘ


[RDF feed](#) 

Nmr:field 50 + 

Nmr:hasPeak [Peak:s4735p1](#) + , [Peak:s4735p2](#) + , [Peak:s4735p3](#) + , [Peak:s4735p4](#) + ,
[Peak:s4735p5](#) + , [Peak:s4735p6](#) + , [Peak:s4735p7](#) + , [Peak:s4735p8](#) + ,
[Peak:s4735p9](#) + , [Peak:s4735p10](#) + , [Peak:s4735p11](#) + , [Peak:s4735p12](#) + ,
[Peak:s4735p13](#) + , [Peak:s4735p14](#) + , [Peak:s4735p15](#) + , and [Peak:s4735p0](#) + 

Nmr:solvent [Chloroform-D1 \(CDCl3\)](#) + 

Nmr:spectrumId 4735 + 

Nmr:spectrumType 13C + 

Nmr:temperature 298 + 



Problem 2: Export RDF in the format it was imported (... not the SMW hard-coded format)



Solution: Storing Original URIs



DrugMet project

Page Discussion

Plastid acetate CoA-transferase complex go:0032283

Facts about Plastid acetate CoA-transferase complex go:0032283

Equivalent URI	http://bio2rdf.org/go:0032283 + 🔍
Label	plastid acetate CoA-transferase complex go:0032283
N0pred:accession	GO:0032283 + 🔍
N0pred:definition	An acetate CoA-transferase complex located in the plastid
N0pred:is a	0009329 + 🔍
N0pred:name	Plastid acetate CoA-transferase complex + 🔍
Ns0pred:sameAs	GO 0032283 + 🔍
Original URI	http://bio2rdf.org/go:0032283 + 🔍
Rights	GO.cite.shtml + 🔍
Type	N0pred:term + 🔍, and N0pred:Term + 🔍
Url	http://bio2rdf.org/html/go:0032283 + 🔍
UrlImage	http://bio2rdf.org/image/go:0032283 + 🔍
XmlUrl	http://bio2rdf.org/xml/go:0032283 + 🔍

UPPSALA UNIVERSITET

Result: *Problem 2 solved!*
Can Query/Export in original RDF format





Special page

Go

Search

SPARQL Endpoint

Enter SPARQL query:

```
@PREFIX w : <http://drugmet.rilspace.org/wiki/Special:URIResolver/> .  
  
SELECT ?p ?o  
WHERE { http://bio2rdf.org/go:0032283 ?p ?o }  
LIMIT 25
```

Query by original URIs: ☒

Output original URIs: ☒

Output format: HTML

Query by Equivalent URIs: ☐

Output Equivalent URIs: ☐

(RDF/XML requires creating triples using [CONSTRUCT](#))

Submit

Navigation

[Main page](#)
[Project Planning](#)
[Recent changes](#)
[Random page](#)
[Help](#)

Semantic tools

[SPARQL Endpoint](#)
[RDF Import](#)

Toolbox

[Special pages](#)

Administration

[ARC2Admin](#)

[Privacy policy](#) [About DrugMet](#) [Disclaimers](#)



So now we have solved the problems:

- Choosing Wiki Titles
- Export in same format as import

... so now we can go ahead and
hook up SMW with an external tool



Talking to SMW from Bioclipse scripts

```
var wikiURL = "http://drugmet.rilspace.org/wiki/";

/* Editing SMW facts */

smw.addTriple("w:Caffeine", "w:is_a", "w:Molecule", wikiURL);
smw.removeTriple("w:Caffeine", "w:is_a", "w:Molecule", wikiURL);

/* Downloading RDF for local querying */

rdfStore = rdf.createInMemoryStore();
rdfStore = smw.getRDF( wikiURL );
result = rdf.sparql( rdfStore,
    "SELECT DISTINCT ?pred WHERE { ?subj ?pred ?obj } LIMIT 10"
)

/* Show some output */

js.print( result );
```



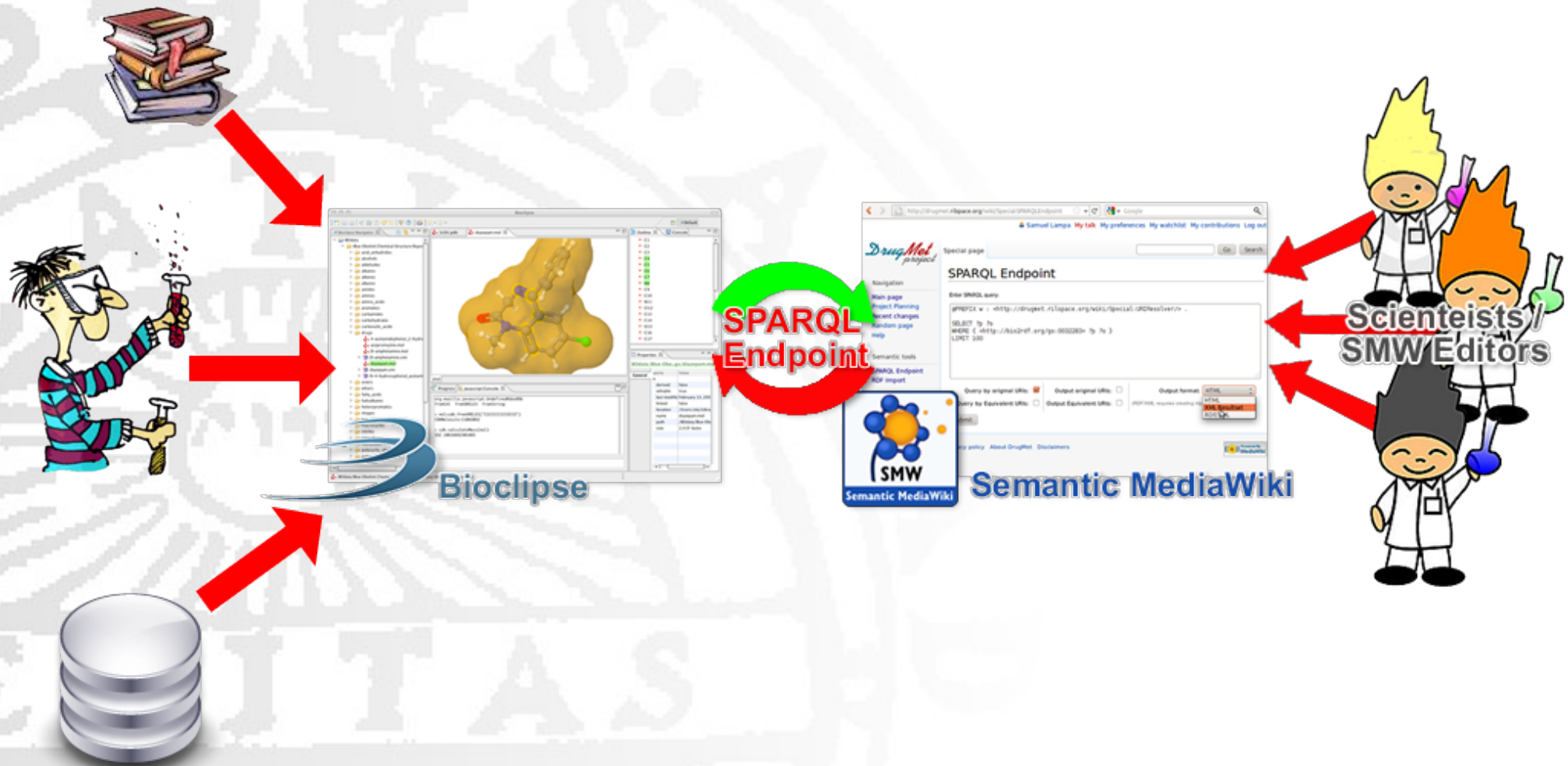
Talking to SMW from Bioclipse (*contd.*)

```
...  
js.print(  
    result  
);
```

```
[["p"],  
["http://www.w3.org/1999/02/22-rdf-syntax-ns#type"],  
["http://www.w3.org/2000/01/rdf-schema#domain"],  
["http://www.w3.org/2000/01/rdf-schema#range"],  
["http://www.w3.org/2000/01/rdf-schema#subPropertyOf"],  
["http://www.w3.org/2000/01/rdf-schema#subClassOf"],  
["http://semantic-mediawiki.org/swivt/1.0#wikiPageModificationDate"], ["http://semantic-mediawiki.org/swivt/1.0#wikiNamespace"], ["http://www.w3.org/2000/01/rdf-schema#isDefinedBy"],  
["http://semantic-mediawiki.org/swivt/1.0#page"],  
["http://www.w3.org/2000/01/rdf-schema#label"]  
]
```



So ... the use case is now possible ...



Some real world usage already

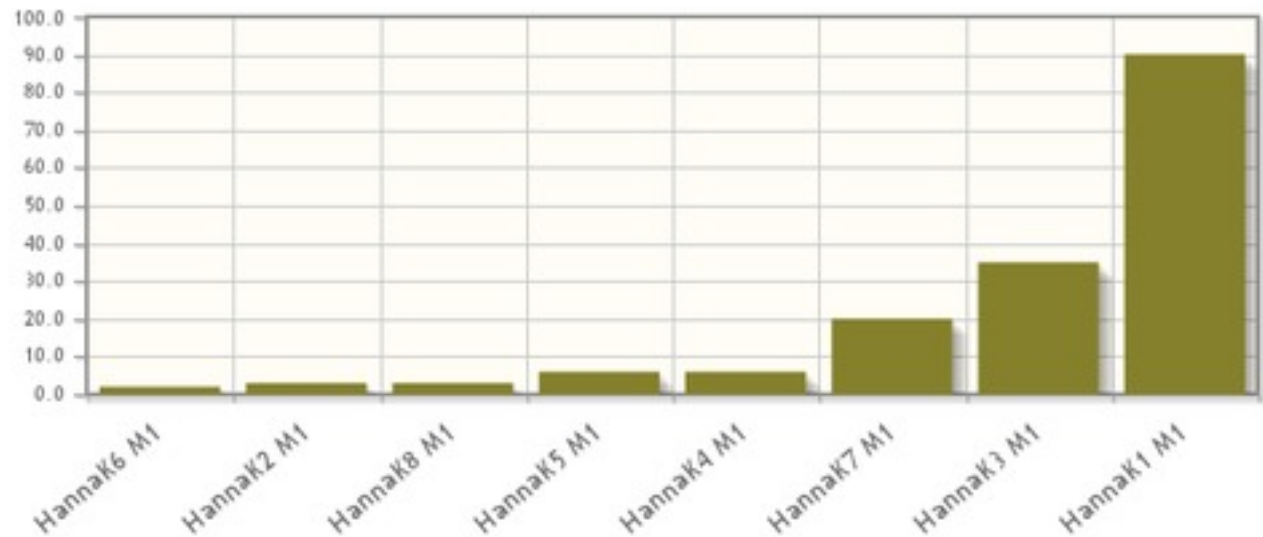


Real world use 1: Visualization of Data

```
{{#ask: [[Category:Measurements]] [[Has Study::{{PAGENAME}}]] [[Has Endpoint::PercentageNonViableCells]]
```

```
| ?Has Endpoint Value  
| format=jqplotbar  
| sort=Has Endpoint Value  
| order=ascending  
| height=250  
| width=600  
}}}
```

Plot



Entity	Dose	Value
Fe3O4	2	
TiO2	3	
carbon nanoparticle	3	



Real world usage 2: Pulling data from R

```
library(rrdf)

endpoint = "http://127.0.0.1/mediawiki/index.php/Special:
SPARQLEndpoint"

query = paste("PREFIX w: ",
"SELECT ?min ?max ?zeta WHERE ",
"{ ?inst a w:Category-3AMetalOxides . ",
"  OPTIONAL { ?inst w:Property-3AHas_Size_Min ?min . }", "
OPTIONAL { ?inst w:Property-3AHas_Size_Max ?max . }", "
OPTIONAL { ?inst w:Property-3AHas_Zeta_potential ?zeta .
}",
"}"
);
```



Real world usage 2: Pulling data from R

```
> data
```

	min	max	zeta
[1,]	15	90	NA
[2,]	15	90	NA
[3,]	15	90	NA
[4,]	15	90	NA
[5,]	15	90	NA
[6,]	15	90	NA
[7,]	15	90	NA
[8,]	15	90	NA
[9,]	15	90	NA
[10,]	15	90	NA
[11,]	15	90	NA
[12,]	15	90	NA
[13,]	10	100	34.2
[14,]	30	60	-17.3
[15,]	20	30	1.8
[16,]	15	90	NA



Outlook

- Proof of concept is there
- Things to fix before useful:
 - SMW 1.6 support
 - Some bugfixes
- Further wishlist:
 - Support more triple stores
 - Skip ARC2 PHP RDF library dependency?
 - SPARQL UPDATE instead of ARC2's SPARQL+
 - "Output as Original URIs" also in RDF Batch export
 - Improve performance



Thank you!

RDFIO Extension: Google Summer of Code project in 2010

Mentor: **Denny Vrandečić** <<http://simia.net/>>

Bioclipse + SMW integration: MSc Thesis project in Bioclipse group, Inst. of Pharmaceutical Biosciences, Uppsala University, Sweden.

Supervisor: **Egon Willighagen** <<http://chem-bla-ics.blogspot.com>>

Samuel Lampa <<http://saml.rilspace.org>>, SMWCon 23 Sep 2011

